

Innovations

Molecular Modeling: An Experimental Tool

The 1920s are sometimes referred to as the "golden age of physics." If recent developments in several other scientific fields are any indication, the golden age of physics may be about to spin off a new golden age of molecular computer modeling. In its broadest sense, molecular modeling is the conjunction of computer technology, crystallography, and theoretical chemistry that focuses on the properties of large biological molecules.

Technology

Technology is rushing headlong to provide faster computers by linking many processors which split the work and avoid, for the time being, the ultimate limit of how fast one can move electrons in wires. Similarly, powerful graphics workstations, the same ones that render *Tyrannosaurus rex* believable in current movies, can now make molecules come alive on a small fluorescent screen in a biochemistry laboratory. The same technology makes it possible via the INTERNET to access, sometimes around the world, databases of sequences or coordinates of proteins and nucleic acids. One database, the Cambridge Database, now stores the relative atomic positions of more than 110,000 small molecules. Another database, the Protein Data Bank at Brookhaven, New York, contains the structures of several hundred large molecules. It has recently become the norm for protein crystallographers to quickly release data on the coordinates of new structures, reversing a long-standing tradition and providing a boon for structural biologists.

Crystallography

Although crystallography (the art of extracting molecular structure from the analysis of scattered X-rays) got its start in the golden age of physics, primarily through the efforts of the father and son team of William and Lawrence Bragg at Cambridge, by 1965 only three protein structures had been solved. In the 1950s, the essential foundation of genetics was provided by Watson and Crick's correct interpretation of the X-ray diffraction data

from fibrous DNA. Since then, improvements in X-ray devices, including the sources and detectors, and in the associated computers and refinement algorithms, have made it possible to sometimes solve the structure (i.e., determine the atomic positions) only weeks after finding a suitable crystal. The increased availability of significant quantities of protein produced with molecular biology techniques has also had a major impact. Once the coordinates that define one structure are determined, variant proteins can be made using molecular biology and their structures can be quickly elucidated, sometimes in a day or two, based on the primary structure.

Theoretical Chemistry

The central icon around which theoretical chemistry clusters is the Schrödinger equation, $H\psi = E\psi$. In this equation, H is an operator containing the second derivative with respect to the positions, ψ is the wave function, and E is energy. Quantum mechanics has as one of its inviolate principles that the wave function (i.e., the solution to the Schrödinger equation), contains all the information that can be measured about a system.

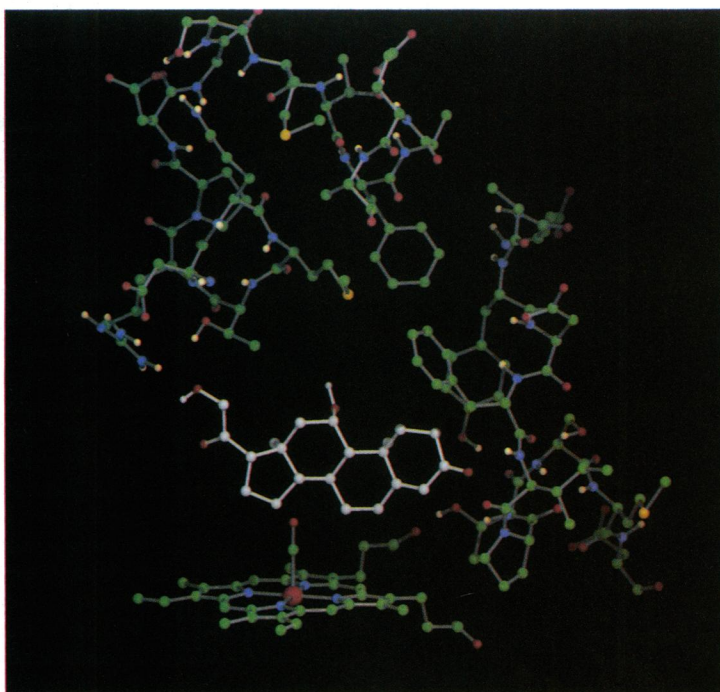
The term "system" refers a part of the universe that can be controlled for study. If we think of the system as a bio-macromole-

cule, along with its environment of other molecules, then the promise of quantum mechanics is clear. We need only solve the Schrödinger equation to know how a molecule will react to stress *in vitro* or *in vivo*. Solving the Schrödinger equation is the major challenge before scientists and they are getting closer. Using these principles, it may one day be possible to study, for instance, the dynamics of a molecular complex important to AIDS.

A major limitation to solving the Schrödinger equation or its derivative equations directly is that computer time requirements grow exponentially with the number of atoms in the system. A new methodology based not on wave functions, the traditional solution of the Schrödinger equation, but on the electron density, is called density functional theory. This new approach derives from a theorem stating that the energy of a system is a unique function of the electron density. Density functional theory is still not on quite as firm a theoretical ground as the conventional methodology; however, because of its potential application to very large systems, it is expected to become a method of choice. Gaussian, Inc., a worldwide supplier of quantum chemical computer code, plans for its next release to include density functional theory options.

Drug Design

The major driving force behind much of the application of these convergent methodologies is the pharmaceutical industry and its unquenchable thirst for new drugs. This industry is responsible for the employment of many scientists with interdisciplinary backgrounds in fields such as synthetic organic chemistry and theoretical chemistry. For example, in its U.S. operations, Glaxo has five molecular modelers, four protein crystallographers, four macromolecular nuclear magnetic resonance spectroscopists, and a specialist in sequence analysis algorithms, with twice this personnel worldwide. Several other major pharmaceutical firms have a similar investment in modeling; in some cases the investment also includes a supercomputer such as a CRAY Y-MP. Much of the research is built around the knowledge of the structure of a key enzyme and a bound substrate or, more



Corticosterone modeled and energy-minimized into the binding pocket of crystallographic P450cam.

likely, a bound inhibitor. Databases and special docking programs can identify other compounds likely to fit into the active-site region. Sometimes new compounds are synthesized to provide a near perfect fit for delivery of a specialized function. More importantly, many more compounds with unknown properties that might have been synthesized are not because the theoretical chemistry criteria (e.g., low space exclusion and complementary charges between ligand and enzyme) are not met. Navia and Murko recently summarized the applications of structural information, including the related impact of molecular modeling, in drug design for hypertension, HIV, emphysema, cancer, ocular hypertension, and coagulation (see Suggested Reading). A recent example of the symbiosis between molecular modeling and directed organic synthesis is the development of inhibitors of thymidylate synthase, an essential enzyme for the cellular production of DNA, which is thus a target for anticancer drugs.

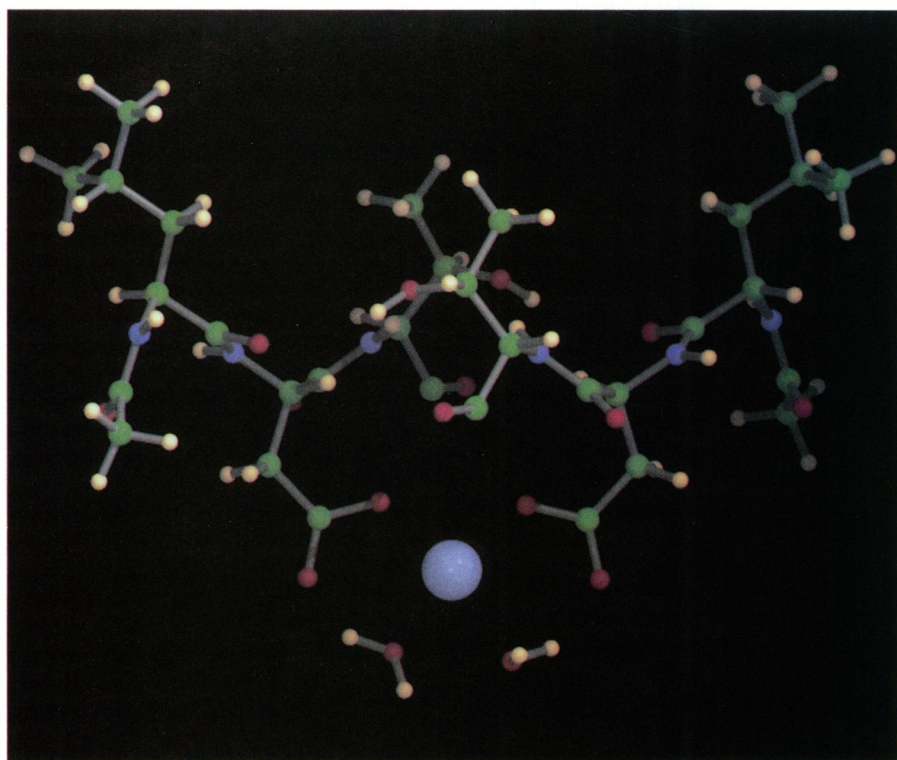
Coincident with the interest of pharmaceutical firms in using computer techniques to design drugs is the evolution of software companies, who develop algorithms and new methodologies to find ligands for sites on macromolecules whose structures are known. One particularly impressive product is the DOCK program developed at the University of California-San Francisco by Kuntz and co-workers.

Toxicology

Developing more effective drugs is one obvious use of the emerging field of molecular modeling. Because drugs and toxic compounds may be viewed as opposite sides of the same coin, all the recent advances in drug development can be applied in toxicology. Scientists at NIEHS have used theoretical techniques to help assimilate a growing body of experimental information derived from molecular biology studies of mammalian P450 enzymes. The NIEHS Laboratory of Reproductive and Developmental Toxicology has shown that the mutation of a few amino acids in mammalian P450 significantly altered the specificity of the enzyme for its substrate. A major roadblock to understanding these results was that the mammalian P450 crystal structure was not yet available. By a combination of molecular modeling, computer graphics, and modified sequence-analysis algorithms, NIEHS scientists demonstrated that the mammalian binding pocket was very likely similar to the bacterial P450 pocket for which a crystal structure is known, and thus we could better understand the specific changes seen in the molecular biology experiments.



H-ras p21 average structure derived from the X-ray structure using the AMBER program.



Zinc ion modeled into the active site of the HIV-1 protease dimer crystal structure.

Dynamics

Many of the current techniques for studying macromolecules involve static structures. How does one introduce the time variable to study the motion, or dynamics, of macromolecular systems? One way is through Newton's equations of motion. Essentially,

one can solve these equations for the motion of all atoms in a molecule if one knows the positions and velocities at one snapshot of time and the potential energy function. This function defines how the atoms interact with one another. The technology of solving Newton's equations for molecules is called

molecular dynamics. To study the motion of a macromolecule for a significant fraction of time (i.e., several hundreds of picoseconds) requires supercomputer-level resources. Scientists at NIEHS have recently used molecular dynamics to study several important enzymatic systems. The *ras* oncogene product p21-H-*ras*, binds guanosine triphosphate (GTP) or guanosine diphosphate (GDP). The active state occurs when GTP is bound, during which time growth signals are sent to the cell. In the GDP-bound state the enzyme is inactive. By using molecular dynamics with the crystal structure used to define a starting position, NIEHS scientists showed that the motion of a hypervariable loop in p21 may be responsible for the activation of the key water-cleaving molecule that deactivates the complex. This finding added support to the crystallographer's speculation that the dynamics of the active site are critical for enzyme function. Molecular dynamics has also been used to investigate the HIV-1 protease system. The catalytic protease of HIV-1 is a dimer, and one of its functions is to precisely clip the large polypeptide generated by the virus. The protease is thus a major target of drug design. Because zinc ions may inhibit the protease, NIEHS scientists studied where a zinc ion might be most likely to bind and then

performed molecular dynamics simulations on the protease with and without the zinc ion bound (see York et al., *EHP* 101: 246-250). The zinc-bound molecular dynamic structure was subsequently used as the foundation for a quantum mechanical calculation to predict the molecular details of a zinc ion bound in the active site.

The Future

The time-dependent Schrödinger equation provides an analog to Newton's equation of motion, but unfortunately it is incredibly difficult to solve, even for small molecules. Theoreticians are currently searching for ways to introduce time dependency into density functional theory so that quantum dynamics can be performed on large systems. To date, the dynamics of molecules consisting of about 100 atoms have been studied in this way. It is likely that the juxtaposition of new developments in computer technology, theoretical chemistry, and crystallography will continue to lead us toward the day when theoretical calculations on large molecules are possible. The deeper understanding of matter that will follow should motivate many new applications in molecular science.

T. A. Darden and L. G. Pedersen

SUGGESTED READING

- Foley CK, Pedersen LG, Charifson PS, Darden TA, Wittinghofer A, Pai EF, Anderson MW. Simulation of the solution structure of the H-*ras* p21-GTP complex. *Biochemistry* 31:4951-4959(1992).
- Iwasaki M, Darden TA, Pedersen LG, Davis DG, Juvonen RO, Sueyoshi T, Negishi M. Engineering mouse P450_{coh} to a novel corticosterone 15- α -hydroxylase and modeling steroid-binding orientation in the substrate pocket. *J Biol Chem* 268:759-762(1993).
- Krieger JH. Computer-aided chemistry poised for major impact on the science. *Chem Eng News* May 11:40-56(1993).
- Navia MA, Murcko MA. Use of structural information in drug design. *Curr Opin Struc Biol* 2:202-210(1992).
- Shoichet BK, Stroud RM, Santi DV, Kuntz ID, Perry KM. Structure-based discovery of inhibitors of thymidylate synthase. *Science* 259:1445-1450(1993).
- Yang W. Direct calculation of electron density in density-functional theory. *Phys Rev Lett* 66:1438-1441(1991).
- Zhang Z, Reardon I, Hui J, O'Connell K, Poorman R, Tomasselli R, Heinrikson R. Zinc inhibition of renin and the protease from human immunodeficiency virus type 1. *Biochemistry* 36:8717-8721(1991).

GIVE FIVE. GOAL THAT REALLY COUNTS

Whatever causes you're giving to now, set a goal to increase your giving to a level that will make a permanent and positive difference. Give Five - 5 hours a week and 5% of your income. The rewards will make you feel like a winner every day of your life. For more information, call 1-800-55-GIVE-5.



Ad Council

A Public Service of The Publication

INDEPENDENT SECTOR

Five hours a week and five percent of your income.